

Understanding Data's Spread

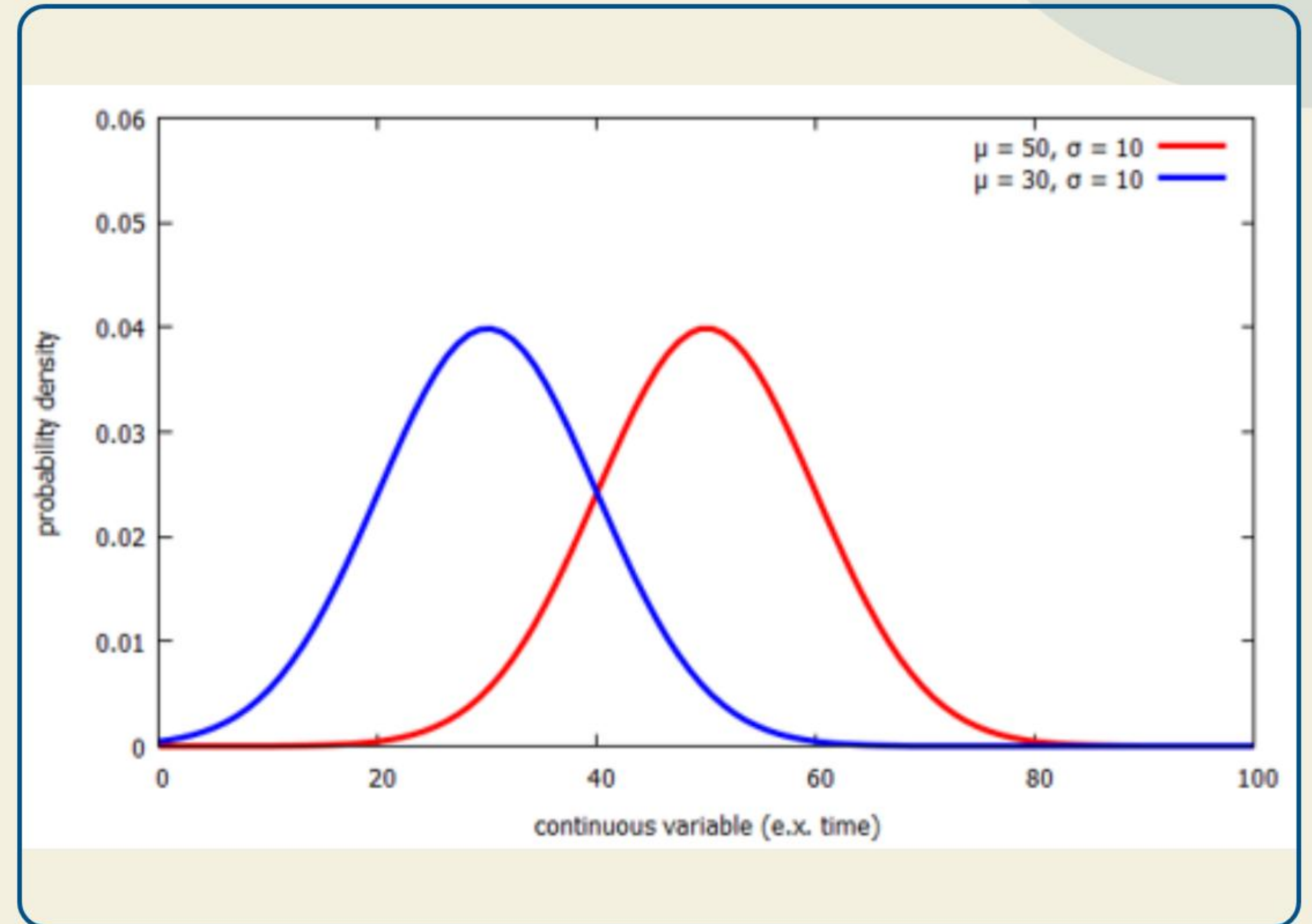
A Guide to Measures of Dispersion

Why Does "Spread" Matter?

Beyond the 'Center'

Central tendency (mean) tells you the center. Dispersion tells you how spread out the data is.

- Two datasets can have the same mean but be vastly different.
- Dispersion measures the variability, consistency, or risk.
- **Example:** Two classes have a 75% mean test score.
 - **Class A:** Scores are 70–80% (low spread).
 - **Class B:** Scores are 50–100% (high spread).



1. The Range

Explaining The Range

80

Range (100 – 20)

The Simplest Measure

The range is the simple difference between the highest and lowest values in a dataset.

- **Pro:** Very easy and quick to calculate.
- **Con:** Extremely sensitive to outliers. It ignores the 'shape' of all other data points.

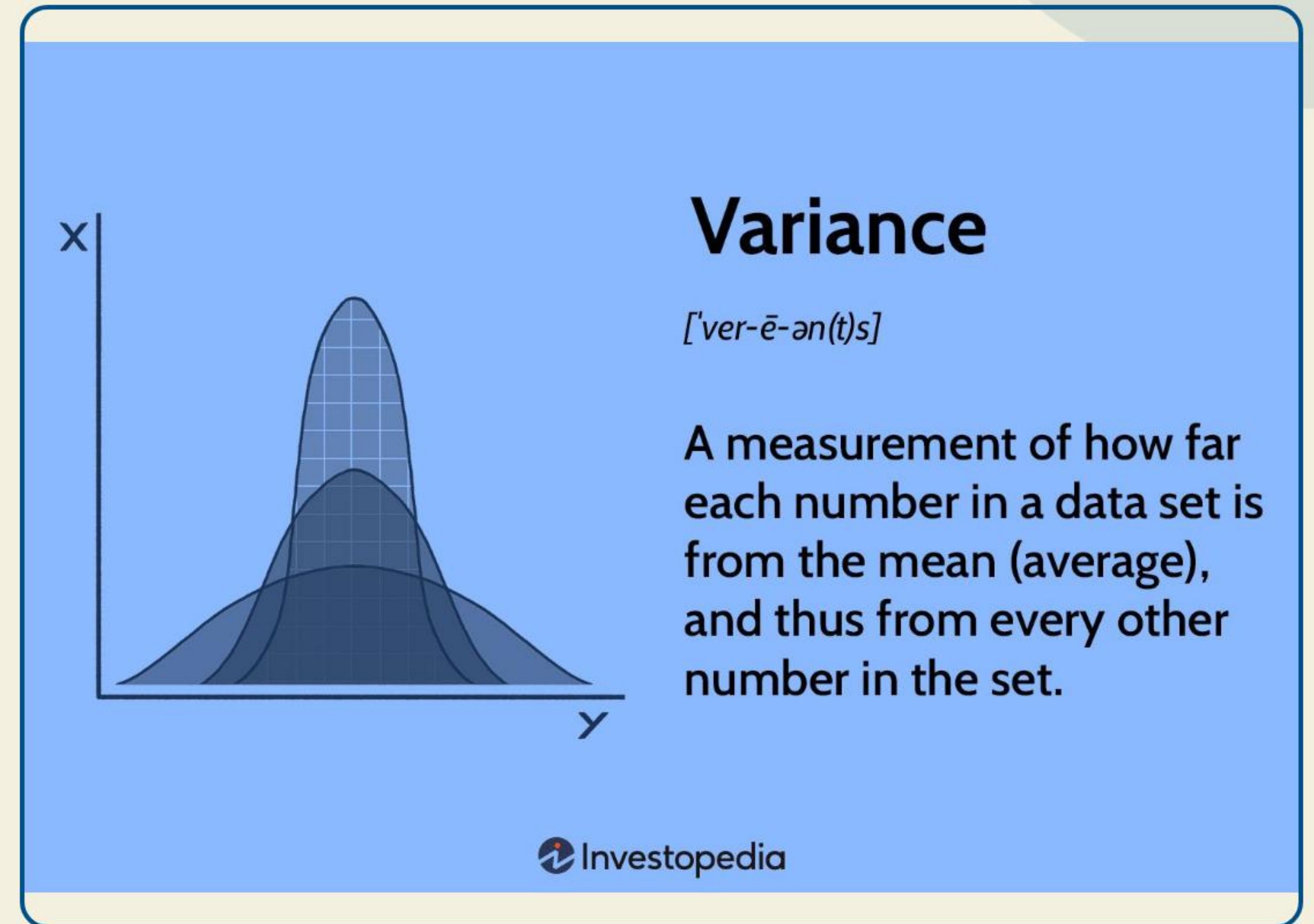
2. The Variance

Explaining The Variance

The 'Average Squared Difference'

Variance measures how far each data point is from the mean, on average.

- It's the average of the **squared** differences from the mean.
- Squaring ensures all differences are positive.
- This gives more weight to outliers.
- **Problem:** The units are **squared** (e.g., $\2 or cm^2), which is hard to interpret.



3. Standard Deviation

Explaining Standard Deviation

The Gold Standard of Spread

The Standard Deviation (SD) is the most common and useful measure of dispersion.

- It's the **square root of the variance**.
- This returns the measure to the **original units** (e.g., back to \$ from \$²).
- A **low SD** means data is clustered near the mean.
- A **high SD** means data is spread out.

VARIANCE AND STANDARD DEVIATION

$$\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2$$

Standard deviation

$$\sigma = \sqrt{\sigma^2}$$

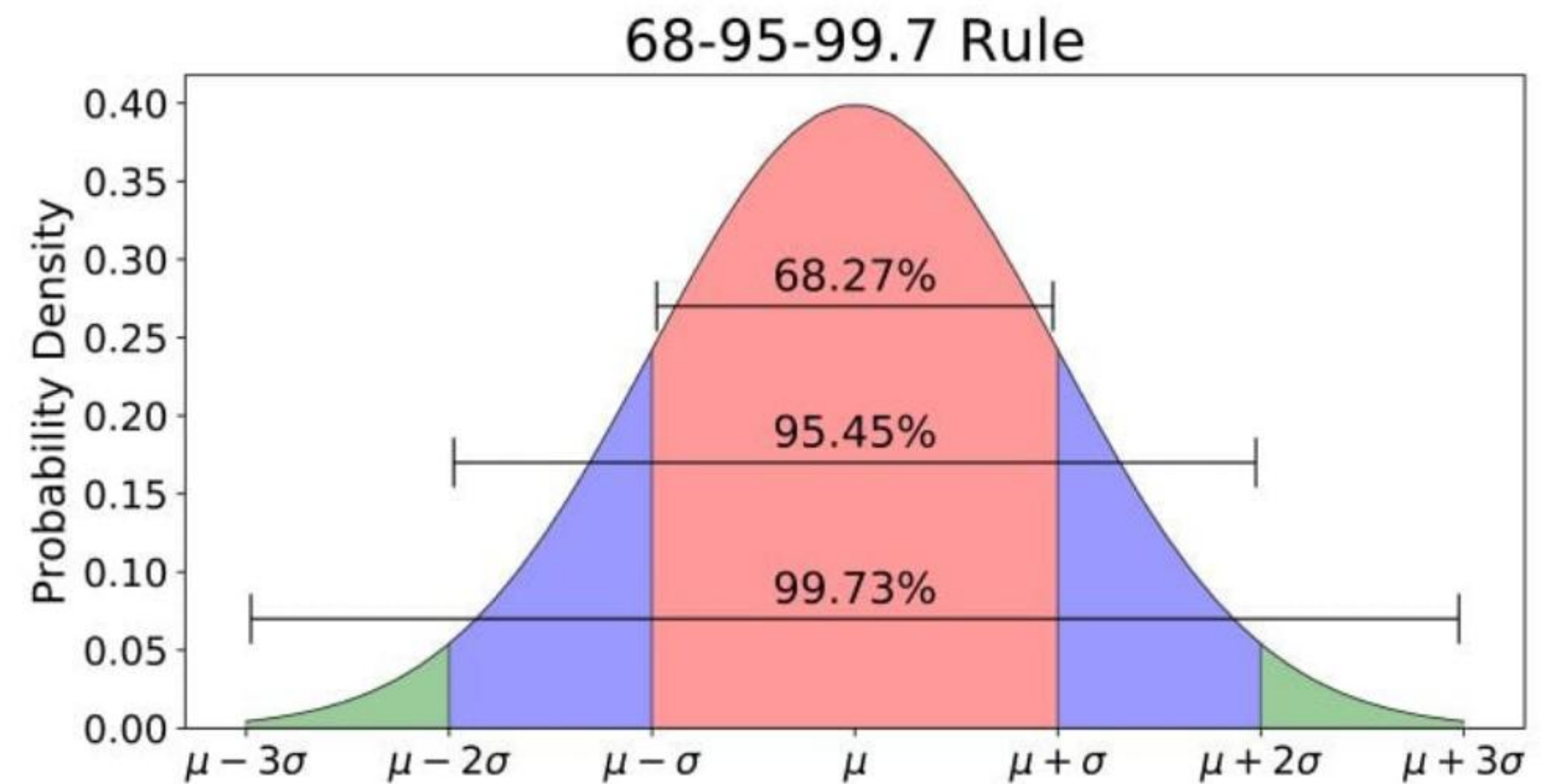
The Empirical Rule (68-95-99.7)

Visualizing Standard Deviation

For normal distributions (bell curves), the SD tells us a lot:

- **~68%** of data falls within 1 SD of the mean.
- **~95%** of data falls within 2 SD of the mean.
- **~99.7%** of data falls within 3 SD of the mean.

This is a fundamental concept in statistics.



Summary of Measures

| Measure | What It Is | Uses Original Units? | Sensitive to Outliers? |
|--------------------|------------------------------|----------------------|---------------------------|
| Range | Highest – Lowest | Yes | Yes (Very) |
| Variance | Avg. Squared Diff. from Mean | No (Squared) | Yes |
| Standard Deviation | Square Root of Variance | Yes | Yes (but less than Range) |

Real-World Applications



Finance

Standard deviation is used to measure **volatility** or risk. A high SD means an investment is riskier.



Science

Used to express the **margin of error** in experiments. A low SD means the results are reliable and repeatable.



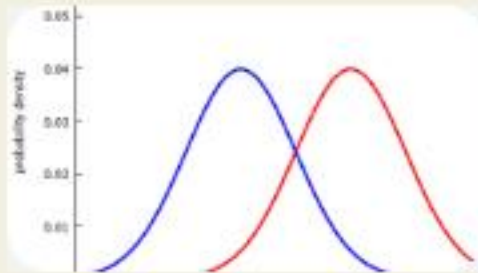
Quality Control

Measures the consistency of a product. A low SD means products are very uniform (e.g., bottle fill levels).

Questions?

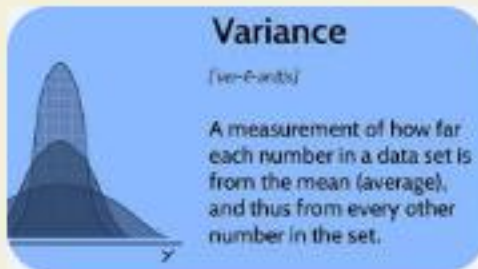
Thank you for your attention.

Image Sources



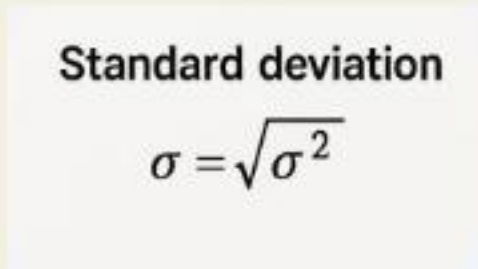
<http://openbooks.library.umass.edu/p132-lab-manual/wp-content/uploads/sites/26/2020/07/Two-normal-distributions.png>

Source: openbooks.library.umass.edu



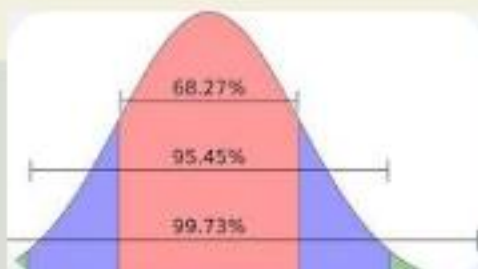
[https://www.investopedia.com/thmb/VDtpekOn1mpNqEm49CZjYmv8zCk=/1500x0/filters:no_upscale\(\):max_bytes\(150000\):strip_icc\(\)/Variance-TAERM-ADD-Source-464952914f77460a8139dbf20e14f0c0.jpg](https://www.investopedia.com/thmb/VDtpekOn1mpNqEm49CZjYmv8zCk=/1500x0/filters:no_upscale():max_bytes(150000):strip_icc()/Variance-TAERM-ADD-Source-464952914f77460a8139dbf20e14f0c0.jpg)

Source: www.investopedia.com



https://miro.medium.com/v2/resize:fit:1024/1*lnRzF1sXDOKWnmm000vTrw.webp

Source: [medium.com](https://miro.medium.com)



https://buitin.com/sites/www.buitin.com/files/styles/ckeditor_optimize/public/inline-images/1_empirical-rule.jpg

Source: buitin.com